

УДК 004

РАЗРАБОТКА И РЕАЛИЗАЦИЯ НЕЙРОСЕТЕВОГО КЛАССИФИКАТОРА НА ПЛИС*В.С. Шумихин**Научный руководитель: Д.С. Григорьев, аспирант каф. ВТ ИК ТПУ
Национальный исследовательский Томский политехнический университет**E-mail: shuhard93@gmail.com*

Abstracts. *Developed a prototype of fragment cascaded neural network with direct serial connections with selectable the activation function of neurons. Fragment executed on programmable logic device (type field-programmable gate array, FPGA) Cyclone 3 (EP3C55) from Altera. It is given a method of designing a fragment of a neural network in the CAD Quartus II 9.1.*

Keywords: neural network, programmable logic device, hardware implementation.

Введение

Одно из основных достоинств нейронных сетей – одновременная обработка большого количества сигналов. Большая часть реализуемых в настоящее время сетей – программная эмуляция, работающая на персональных компьютерах и специализированных серверах. Помимо достоинств, таких как простота, достигаемая программной эмуляцией, и повсеместная доступность персональных компьютеров, эти станции имеют и недостатки, например, избыточность и большую потребляемую мощность.

По этой причине возникает необходимость разработки специализированного аппаратного ускорителя нейросетевого алгоритма. Ввиду узкой специализации нейросетевого алгоритма, его аппаратная реализация в виде заказной СБИС не требует организации серийного производства и достаточно дорога. Отсюда следует, что наиболее эффективной аппаратной реализацией нейронной сети является реализация на программируемых логических интегральных схемах – ПЛИС (ведущие изготовители – Xilinx, Altera).

Структура фрагмента нейронной сети

Ввиду того, что размещение всей нейронной сети на кристалле невозможно по причине ограниченности физических ресурсов кристалла, возникает необходимость разбивать сеть на каскадируемые фрагменты. Для реализации был принят фрагмент нейронной сети с прямыми последовательными связями, как наиболее часто встречающийся в реальных задачах, состоящий из 4-х нейронов. Каждый нейрон имеет 4-е входа. Был выбран способ вертикальный каскадирования с увеличением числа нейронов в слое.

Каскадирование может быть как физическим (каждый фрагмент располагается на отдельном кристалле), так и виртуальным (фрагмент зацикливается на одном кристалле), или смешанным. На выбор одного из указанных способов влияет заданное время решения задачи. При физическом каскадировании время решения задачи меньше времени решения при виртуальном каскадировании в число раз примерно равное числу физически реализованных фрагментов.

Структура нейрона

Ниже предлагается схема замещения искусственного нейрона, разбитая на четыре функциональных блока: умножители, память, сумматор, блок функции активации – далее используемая в ПЛИС с сохранением этих функциональных блоков.

Входные данные поступают на умножители, которые реализуют операцию умножения входных сигналов на веса синоптических связей, которые хранятся в памяти блока RAM. Веса являются константами. Входные сигналы и веса связей являются числами с плавающей точкой одинарной точности. Сумматор последовательно складывает полученные ранее про-

изведения, и подает сумму в блок функции активации. Блок функции активации отвечает за выбор функции активации (пороговая или сигмоидальная) и расчет выходного сигнала по выбранной функции.

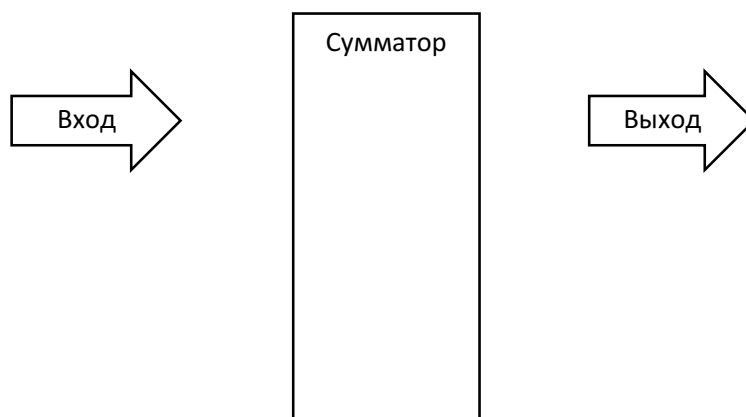


Рис. 1. Схема замещения нейрона

Блок функции активации может быть размещен вне нейрона и быть общим для всего фрагмента, но тогда теряется параллельная обработка сигналов нейронами, т. к. каждый нейрон будет обращаться в блок функции активации и ожидать пока блок закончит работу с предыдущим нейроном. Следовательно, такое решение ведет к потере быстродействия всего фрагмента и каскада фрагментов. Поэтому блок функции активации был размещен в нейроне.

Заключение

Разработан прототип каскадируемого фрагмента нейронной сети с прямыми последовательными связями с возможностью выбора функции активации нейронов. Максимальная тактовая частота работы схемы 90 МГц. Время вычисления выходного вектора по входному: 70 нс. Разработка фрагмента осуществлялась в САПР Quartus II 9.1 на языке VHDL.

В ближайшее время предполагается провести более детальные временные и топологические оптимизации проекта с целью повышения быстродействия фрагмента. Необходима возможность настройки весов связей. Необходима также разработка специализированных библиотечных нейроэлементов (сумматоров, блоков функции активации), с помощью которых предполагается увеличить производительность и функциональные возможности нейросетевого устройства.

Список литературы

1. Neural Network Implementation Using FPGA: Issues and Application / A. Muthuramalingam, S. Himavathi, E. Srinivasan. International Journal of Information and Communication Engineering 4:6, 2008.
2. Neural network training based on FPGA with floating point number format and it's performance / Mehmet Ali Cavuslu, Cihan Karakuzu, Suhap Sahin, Mehmet Yakut. Neural Comput & Applic 20:195, 2011.
3. Галушкин А.И., Кирсанов Д.В., Цифровые нейрочипы специализированные цифровые БИС для нейрокомпьютеров // Зарубежная радиоэлектроника. – 1999. – № 1. – С. 17–37.
4. Хайкин С., Нейронные сети / Вильямс 2006.